Patient sharing and population genetic structure of methicillin-resistant Staphylococcus aureus

Patient sharing and population genetic structure of methicillin-resistant *Staphylococcus aureus*

Weixiong Ke[a], Susan S. Huang[b], Lyndsey O. Hudson[c], Kristen R. Elkins[b], Christopher C. Nguyen[b], Brian G. Spratt[c], Courtney R. Murphy[b], Taliser R. Avery[d], and Marc Lipsitch[a,e,1]

[a]Department of Epidemiology and [e]Center for Communicable Disease Dynamics and Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115; [b]Division of Infectious Diseases and Health Policy Research Institute, Irvine School of Medicine, University of California, Irvine, CA 92617; [c]Department of Infectious Disease Epidemiology, Imperial College London, London W2 1PG, United Kingdom; and [d]Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA 02215

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved February 16, 2012 (received for review August 26, 2011)

Rates of hospital-acquired infections, specifically methicillin-resistant *Staphylococcus aureus* (MRSA), are increasingly being used as indicators for quality of hospital hygiene. There has been much effort on understanding the transmission process at the hospital level; however, interhospital population-based transmission remains poorly defined. We evaluated whether the proportion of shared patients between hospitals was correlated with genetic similarity of MRSA strains from those hospitals. Using data collected from 30 of 32 hospitals in Orange County, California, multivariate linear regression showed that for each twofold increase in the proportion of patients shared between 2 hospitals, there was a 7.7% reduction in genetic heterogeneity between the hospitals' MRSA populations (permutation *P* value = 0.0356). Pairs of hospitals that both served adults had more similar MRSA populations than pairs including a pediatric hospital. These findings suggest that concerted efforts among hospitals that share large numbers of patients may be synergistic to prevent MRSA transmission.

Methicillin-resistant *Staphylococcus aureus* (MRSA), one of the most common and virulent nosocomial pathogens, is also an increasingly important cause of community-acquired disease. MRSA strains, particularly those associated with hospitals, are often resistant to multiple antibiotics, limiting treatment options. *S. aureus* is carried asymptomatically in ∼30% of healthy adults and is shown to be a major cause of invasive disease among hospitalized patients (1); MRSA makes up a growing proportion of nosocomial *S. aureus* infections in many countries. The circulation of a small number of MRSA clones that characterizes the current epidemic is thought to be mainly the result of between-patient transmission rather than de novo appearance of resistance in patients exposed to antibiotics (2), because the appearance of a new MRSA strain requires acquisition of a *mec* resistance element, a relatively rare event. The level of genetic variation occurring in *S. aureus* within identifiable clonal lineages allows the use of genetic markers to track transmission of these lineages and sublineages (3, 4).

The prevalence of MRSA varies considerably both within and between countries (5, 6). About 30% of the *S. aureus* causing bloodstream infections is methicillin resistant in the United Kingdom, whereas that proportion is ∼1% in The Netherlands and Scandinavian countries (7). Among countries with high endemic MRSA infection rates, the proportion is highest in large teaching hospitals (6, 8), where the highest frequency of new emerging MRSA clones has also been reported (9–12). The proposed reasons include increased antibiotic use and increased prevalence of medical procedures and serious medical conditions associated with MRSA acquisition and disease (13). Because MRSA can be carried asymptomatically for a long time (14), readmission could introduce a previously acquired strain into a new hospital (15). Thus, failure of one hospital's infection control could in principle affect the prevalence of MRSA in other hospitals that share patients with it (16). Previous studies have suggested that patient transfer or patient referral patterns (17) could affect the prevalence of MRSA in hospitals (1, 2, 16, 18, 19), on the basis of theoretical arguments and observations that clones of MRSA appear in neighboring hospitals. Population genetics can provide a test of the hypothesis that patient sharing plays an important role in MRSA dynamics: If so, one might expect that hospitals that share large numbers of patients would also tend to share genetically similar populations of MRSA.

In the current study, we sought to investigate whether the pattern of genetic relatedness among MRSA isolates from hospitals within Orange County (OC), California, was consistent with a significant role for patient sharing in determining the population of MRSA strains within a hospital. OC is well suited for this study because it is the fifth largest county in the United States, and it has relatively low population flow from three of its four sides. A finding that MRSA isolates from hospitals that share more patients tend to be related to one another would provide an independent line of evidence for the importance of patient transfer in spreading MRSA from hospital to hospital. For *S. aureus* genotyping, we used the *spa* locus, which encodes protein A, a species-specific protein known for its IgG binding capacity (3). This locus features highly polymorphic internal regions due to short tandem repeats (STRs) (20) and therefore serves as a good target for molecular genotyping (*spa* typing). This genotyping method has been demonstrated to be useful in researching transmission, outbreaks, or geographic distributions (3, 21, 22). Using *spa* typing, we sought evidence on how MRSA strains "travel" with patient flow, to infer how patient transfer might influence MRSA spread among hospitals.

Results

Summary of Overall Approach. We used Wright's F statistics to measure genetic heterogeneity between MRSA populations in hospitals and groups of hospitals and used pairwise regression analysis supplemented by group-level analysis as our methods, as described in *Materials and Methods*.

In the pairwise regression analysis, we calculated the heterozygosity of each pair of hospitals (H_R) (for each pair we regard the two hospitals as a group) and the heterozygosity of each single hospital within a pair (H_S). F_{SR} was calculated for each pair of hospitals accordingly and was used as our response variable to measure the genetic dissimilarity between a pair of hospitals. A positive coefficient for a predictor of F_{SR} indicates that hospitals are more divergent from one another, whereas a negative coefficient indicates that hospitals are more similar to one another.

Author contributions: W.K., S.S.H., B.G.S., and M.L. designed research; W.K., C.R.M., and M.L. performed research; L.O.H. contributed new reagents/analytic tools; W.K., L.O.H., K.R.E., C.C.N., C.R.M., and T.R.A. analyzed data; and W.K., S.S.H., B.G.S., C.R.M., and M.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

See Commentary on page 6364.

[1]To whom correspondence should be addressed. E-mail: mlipsitc@hsph.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1113578109/-/DCSupplemental.

The distribution of F_{SR} is shown in Fig. S1*A*. We found that the distribution of this pairwise F_{SR} is significantly skewed, so we log-transformed this variable in the regression analysis, and the new distribution is shown in Fig. S1*B*. This procedure was also carried out for the main predictor variable, patient flow, as shown in Fig. S1 *C* and *D*.

In group-level analysis, we used the F_{ST} statistics calculated from heterozygosity of all 30 hospitals (H_T) and heterozygosity of each individual hospital (H_S) to measure heterogeneity or the reduction of heterozygosity from the total population level to the individual hospital level. For this study, F_{ST} = 0.0853, and H_T and H_S from which it is calculated = 0.719 and 0.658, respectively. As F_{RT} measures the reduction of heterozygosity when hospitals are grouped compared with the total 30 hospitals, this result implies that the best possible grouping (that is, each single hospital is viewed as one group) could do no better than to achieve an F_{RT} = 0.0853 ($F_{RT} = F_{ST}$).

Predictors of Similarity Between Pairs of Hospitals. ***Individual categorizing variables.*** Table S1 shows the characteristics of all 30 hospitals. For pairwise analyses described below, dichotomous variables for private insurance proportion, Medicaid coverage, hospital size, and proportion Hispanic were created to reflect above-cut point or below-cut point values.

Pairwise analysis of similarity. For each of the 435 hospital pairs, we calculated and log-transformed F_{SR} to get log-F_{SR}, serving as a measure of population differentiation between them. We performed multivariate regression on the relationship of log-F_{SR} to log-transformed patient flow and to pairwise geographic distances between hospitals, average isolates collected, and the dichotomous variables mentioned above. The *P* values presented in all of our regression analyses are multiple Mantel test permutation *P* values, used to account for dependency between observations involving pairs of hospitals. The results are presented in Table 1. Greater patient flow between a pair of hospitals (log flow) was associated with reduced pairwise F_{SR}, i.e., greater similarity in MRSA populations between hospitals (coefficient = –0.115, *P* value = 0.0356). Another predictor of similarity in MRSA populations was for both to be nonpediatric (ped00, coefficient = –0.801, *P* value = 0.0448). Univariate analyses are shown in Table S2.

In addition, we calculated Pearson correlation coefficients for each pair of the 13 variables to assess collinearity of predictors. We found relatively large correlation between the proportion of Hispanic patients and Medicaid coverage (coefficient for ethnicity11 and medicaid11 = 0.581 and for ethnicity00 and medicaid00 = 0.564), and between the average number of isolates collected and hospital size (coefficient for SSize and size11 = 0.616 and for SSize and size00 = –0.680). This result suggests that these variables might have had small and insignificant effects in our models due to collinearity. We thus performed multivariate analyses with each (group) of these variables removed in turn, as shown in Tables S3–S6. The results of these models were consistent with those in the primary analysis. In these alternative models, the coefficient of log flow ranged from –0.110 to –0.134 (vs. –0.115 in the base model) and remained statistically significant. Exclusion of sample size led to a statistically significant increase in similarity between pairs of large hospitals, compared with pairs containing a small and a large hospital (Table S3). Exclusion of hospital size variables led to a statistically significant association between larger sample size and greater similarity between the hospitals (Table S4). Neither exclusion of the ethnicity variables (Table S5) nor exclusion of the Medicaid variables (Table S6) produced a statistically significant association with the other, but with Medicaid excluded there was a trend toward greater similarity of hospitals having >20% Hispanic patients.

Although the correlation between patient flow—our main variable of interest—and geographical distance was not as large (coefficient for log flow and dist = –0.378), we conducted another multivariate analysis with distance excluded, as shown in Table S7. The result suggested that by removing the distance predictor, a slightly greater degree of similarity was associated with greater patient flow (coefficient = –0.128).

Two of the hospitals have fewer isolates available (hospital 21, six isolates; hospital 29, four isolates). To verify that our results were not driven by them, we removed them from the dataset and ran the full multivariate regression again. The coefficient on patient flow was almost unchanged (–0.114) although the Mantel *P* value increased to 0.053 (Table S8).

As we hypothesized, our results based on the full multivariate model showed that hospitals sharing more patients have significantly more similar MRSA populations, after adjustment for other possible confounders. For each factor of 2 increase in patient flow (see *Materials and Methods* for definition), there is an associated $1–2^{-0.115}$ = 7.7% reduction in pairwise F_{SR} between the hospitals, whereas the interquartile range of hospital pairs for log flow was –11.51 to –8.74, which corresponds to a 19.8% reduction in pairwise F_{SR}. Another variable was also statistically associated with increased similarity of MRSA populations: Hospitals that both served adult patients tended to have populations that were more similar to one another. Visual inspection suggested that none of these results were driven by individual outliers.

Predictors of Similarity at the Group Level. ***Estimating grouping efficiency.*** We used group-level analysis to supplement our results in pairwise analysis. This method divides hospitals into groups by a given criterion and calculates F_{RT} for the grouping. F_{RT} here measures the reduction of heterozygosity caused by grouping relative to all hospitals without grouping and serves as a mea-

Table 1. Results of pairwise multivariate analysis

Variable no.	Variable name	Description	Coefficient	*P* value*
1	Log_flow	Log-transformed pairwise patient flow	–0.115	**0.0356**
2	SSize	Average sample size of MRSA isolates provided by a pair of hospitals	–0.010	0.2474
3	Medicaid11	Indicator of both hospitals having Medicaid >10%	–0.342	0.2476
4	Medicaid00	Indicator of both hospitals having Medicaid no more than 10%	–0.191	0.5510
5	Private11	Indicator of both hospitals having private insurance >35%	0.041	0.8742
6	Private00	Indicator of both hospitals having private insurance no more than 35%	0.081	0.7068
7	Size11	Indicator of both hospitals having annual admission >10,000	–0.133	0.7198
8	Size00	Indicator of both hospitals having annual admission no more than 10,000	–0.026	0.9404
9	Ethnicity11	Indicator of both hospitals having >20% Hispanic patients	–0.077	0.8102
10	Ethnicity00	Indicator of both hospitals having no more than 20% Hispanic patients	–0.166	0.5714
11	Ped00	Indicator of both hospitals being mainly nonpediatric	–0.801	**0.0448**
12	Ped11	Indicator of both hospitals being pediatric	0.368	0.5022
13	Dist	Distance between a pair of hospitals in kilometers	0.004	0.7066

*Permutation *P* values were calculated by multiple Mantel test permutation. *P* values that reached significance are in boldface type.

surement of between-group heterogeneity—that is, a meaningful grouping should give a higher F_{RT}. In group analysis, "T" refers all 30 hospitals (MRSA population), "R" refers to a group of hospitals, and "S" refers to a single hospital.

We first show the evaluation criteria for groupings. In Fig. 1, the red line gives the theoretical best F_{RT} any grouping can achieve, and the "X"s and error bars show how well random groupings can perform. We used a genetic algorithm to search for groupings with nearly the best possible results attainable for a given number of groups, and we constructed random groups to give the null distribution of F_{RT}; see *Materials and Methods* for details of these approaches.

Clearly, with more groups it is possible to obtain a higher value of F_{RT}; in the limiting case of 30 groups each representing one hospital, we would have had $F_{RT} = F_{ST}$. Thus, we want a grouping scheme to have high F_{RT} while also having a relatively low number of groups.

Fig. 1 also gives evaluations of grouping by a few individual criteria (corresponding to some of the predictors in pairwise analysis). Roughly, these results agree with that from the pairwise analysis. Details can be found in Table S9, and the result of network grouping is described below.

Grouping by the information theoretic approach (network grouping). This method creates groups of hospitals such that there is more frequent patient sharing within each group than between groups. The method used to create groups within the patient-sharing network uses an information theoretic approach (23) that creates groups on the basis of minimizing the expected description length of an idealized random walk on the network; for details see ref. 23. The four groups are illustrated in Fig. S3.

Given this grouping scheme, the H_R, or the heterozygosity on groups, is found to be 0.705, and we calculated $F_{RT} = 0.0204$ (Fig. 1), with P value = 0.050. At four groups of hospitals, the maximum F_{RT} by genetic algorithm is 0.0448, and thus the grouping is achieving 45.5% of the best achieved by the GA.

Discussion

This study used patient sharing data together with *spa* genotyping of MRSA strains to analyze how genetic similarity of MRSA depends on patient sharing networks in Orange County hospitals.

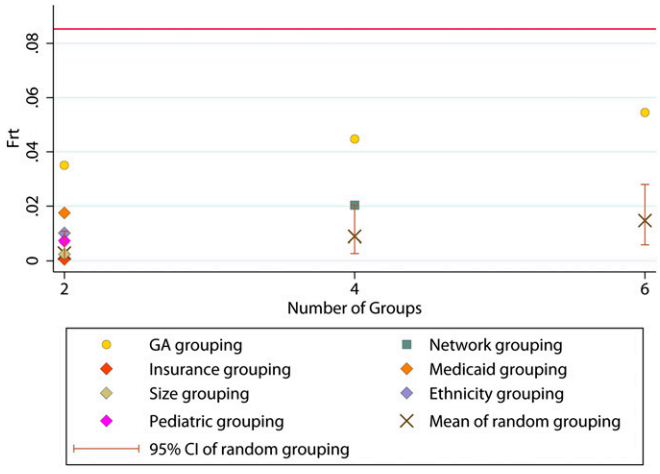

Fig. 1. F_{RT}s by various grouping methods. The uppermost horizontal line represents the maximum value F_{RT} can achieve under any condition (which is F_{ST} of the total population of 30 hospitals). F_{RT}s attained by the genetic algorithm (GA) serve as an estimate of the largest F_{RT} attainable in practice by numerical optimization. The mean and 95% confidence interval (CI) of random grouping display the distribution of randomly generated F_{RT}s at a given number of groups, serving as an estimation of "background". Other symbols of various colors stand for F_{RT}s of different groupings, the detailed description of which can be found in the main text.

In the United States, patient sharing is driven by both the patient and the health provider (24). Patients may choose services at different locations and are influenced by many factors, including their insurance policy, which may restrict patients' choice of hospitals. Changes in insurance policies, transfers to more advanced hospitals for better care, and other reasons might cause patient moves between hospitals. Many theoretical studies have addressed the association between MRSA prevalence and patient referral between hospitals (for example, see refs. 2 and 15), and a positive correlation has been predicted. We address this question from a bacterial population genetic perspective (25, 26), using systematic samples from hospitals within a single county.

We found that the extent of patient sharing between hospitals predicts the extent of genetic similarity between isolates of MRSA obtained from them. Using both pairwise and group-level analysis, we found that the more patients were shared between hospitals, the more similar their MRSA appeared at the *spa* locus. Regression analysis for pairs of hospitals showed significantly more similarity between MRSA from pairs of hospitals that shared more patients, after accounting for other potential predictors including physical distance. Meanwhile, our group-level analysis found that the grouping that classified hospitals on the basis of patient sharing gave an F_{RT} that was significantly better than that of randomly generated grouping schemes.

Agent-based models using these data have found that outbreaks in one hospital could translate to increases in MRSA burden in another hospital (27). The finding that greater patient sharing is associated with greater genetic similarity of MRSA strains, after adjusting for possible confounders, supports the idea that patients track contagious pathogens across hospitals. This result is important given the perhaps unexpectedly large volume of patient sharing that occurs during routine medical care in US hospitals (17).

Patients in Orange County tend to be admitted to hospitals close to their homes. As a result, it is likely that similarities in MRSA strains found in patients who reside near one another could be caused by shared exposure to the same hospitals, as well as by transmission within the community. The finding that hospitals caring for adult patients had more similar MRSA strains than pediatric hospitals may be a further indicator (beyond our findings about patient flow) that MRSA genotypes segregate with patient sharing patterns, since pediatric and adult medical care is segregated in the United States. More definitive studies showing reduction in MRSA burden and strain similarities following regional hospital collaboratives are needed to further understand the contagious impact of sharing patients and the magnitude of prevention that is achievable.

To date, despite a number of theoretical studies suggesting the possible benefits of interventions coordinated among groups of hospitals sharing patients and the possible "externalities" of high MRSA rates in one hospital increasing those in neighboring hospitals, policies such as Medicare reimbursement treat MRSA infections as a problem of the individual hospital, with the effect, as has been argued, that "current Medicare rules subsidize MRSA pollution" (ref. 28, p. 163–182). A possible reason for this seeming disconnect between modeling evidence and policy is the lack of direct empirical evidence that populations of MRSA in one hospital can be traced to sharing of patients from other hospitals at the local level. Many prior empirical studies have documented the spread of clones between hospitals, regions, or countries or have shown that individuals with MRSA colonization are transferred between hospitals. Other studies, which showed that referral hospitals had the highest rates of MRSA infection, did not disentangle whether this association was due to greater numbers of transferred patients, sicker patients, or other factors. This study provides rigorous evidence for the role of patient sharing within a local area in leading to measurable changes in the MRSA population in individual hospitals over a sustained period; moreover, the genetic

evidence provided here is an independent line of evidence that confirms the importance of patient sharing. This finding implies that there is hope for synergistic impact to reduce MRSA, with concerted efforts by hospitals to implement prevention strategies together. Had we found that MRSA strains were indiscriminately found throughout all hospitals, this result would have suggested that a county-wide approach to MRSA containment would be necessary. Instead, it appears possible that targeted approaches might well produce substantial impact when applied to a small group of hospitals that are strongly connected by patient sharing.

There are several limitations to our study. First, our measure of patient sharing considered transfer of patients between hospitals, in both directions and regardless of their MRSA colonization status. A more directly relevant measure, if it were available, would be the transfer of patients colonized with MRSA from each hospital to the other (19). On the other hand, the demonstration here that overall patient transfer is a predictor of MRSA similarity between hospitals suggests that overall patient transfer is an adequate surrogate for the effect of sharing of MRSA-colonized patients, at least for analyses of this type. A related limitation is that whereas 92% of adults in the data set for patient transfer had identifiers (thus only 8% were untracked), the majority (86%) of children lacked such identifiers. However, of these, 63% were <6 mo of age, and a large proportion of these would have been hospitalized at birth and would not have been readmitted (17). Second, the current study did not assess the impact of strains categorized as hospital-onset (HO)-MRSA vs. community-onset (CO)-MRSA (3). We did not assess this distinction partly because community and healthcare reservoirs are mixing and because, at a hospital level, there were often too few strains to make these types of evaluations. Moreover, all strains were treated equally, assessing only whether they were distinguishable by our typing method. Third, patient sharing data could be used in different ways that take into account patient sharing directedness, time of transfer, and length of stay in hospitals, which might provide other interesting findings but were not implemented due to data limitations. Fourth, unmeasured or residual confounding of the association between patient flow and MRSA population similarity is a possibility, as in all such studies. A potential confounder of particular concern is that hospitals that transfer many patients may also draw from the same patient population, so that similarity of the catchment populations leads to importation of similar strains, which could explain MRSA population similarities independent of any causal effect of patient sharing. To address this problem, we included in our model a variable for distance between the two hospitals; compared with a model omitting distance (Table S7), the baseline model (Table 1) had a similar, but slightly smaller effect of patient sharing. Moreover, distance between hospitals was not a significant predictor in the baseline model, nor was it significant in univariate analysis, whereas patient sharing was. In addition, we included variables for shared demographic characteristics of patients, to further eliminate spurious associations with patient sharing that are in fact caused by similar patient populations. Nonetheless, because none of these variables perfectly captures similarities in patient populations, it is possible that the association between patient sharing and genetic similarity of MRSA remains biased by some of these factors. Specific *spa* types associated with pediatric and adult patients in this population have been described recently (29). Fifth, although the *spa* genotyping method used in the current study is widely used as a fast and reliable genotyping technique for *S. aureus*, we might obtain more meaningful results if we used higher-resolution typing systems. Finally, we noticed that some of our hospitals have relatively fewer isolates available. Although the exclusion of these hospitals did not qualitatively affect the results of our pairwise analysis, we found that removing hospitals 21 and 29 made our network grouping result insignificant. The main reason for this loss of statistical significance is that each hospital has a *spa* type that is rare among all hospitals (appearing only in the hospitals that are in the same networking group), and its removal, combined with the fact that these hospitals have fewer isolates, made the distribution of random F_{RT}s generated by random grouping higher. Although we preserved these hospitals in group-level analysis as we believe these rare isolates indicated within-group similarity, more isolates from these hospitals, if possible, are strongly desired.

In summary, we found that patient sharing patterns across hospitals are likely to be correlated with MRSA genetic heterogeneity, along with several other hospital characteristics. This study is a unique regional analysis of a relatively enclosed large metropolitan region of 3 million people. It performs a comprehensive analysis of whether hospitals that share patients also share MRSA strains. It provides evidence of local ecosystems within a single region that are associated with shared patients and suggests that certain groups of local hospitals could make concerted and synergistic efforts to reduce the prevalence of important resistant pathogens and reduce healthcare-associated disease.

Materials and Methods

Study. We conducted a population-based, prospective collection of clinical isolates of MRSA from 30 of 32 hospitals in OC, California as described elsewhere (17). The geographical distribution of these hospitals is shown in Fig. S4. This study was approved by the Institutional Review Board of the University of California Regents.

Isolate collection, specimen- and hospital-level data, and laboratory methods are described in *SI Text*.

Measuring the Genetic Similarity in MRSA Between Pairs or Groups of Hospitals. We adopted a standard measure of genetic similarity: Wright's *F* statistics (30). *F* statistics detect population substructure measured by a given genetic locus of interest. One first calculates the "heterozygosities" of this locus on different levels; here, heterozygosity corresponds to the probability that two randomly chosen isolates will differ at the locus of interest. Three hierarchical levels of population were used in this study: (*i*) subpopulations (S) refer to the bacteria isolated from a single hospital, (*ii*) "regions" (R) refer to the bacteria from a subset (group) of hospitals less than the 30 total hospitals in our study, and (*iii*) the total population (T) refers to all bacteria included in our study. Note that in pairwise analysis, we only have R (a pair of hospitals) and S (a single hospital)—T is not used in the pairwise analysis. If there is any population substructure, then the heterozygosity calculated for the total population will be higher than the weighted average of that calculated for each group individually. In this report, we use the term "heterogeneity" to refer to high *F* statistics implying genetic differentiation between different populations.

Formally, heterozygosity of a population (in terms of one genetic locus) is defined as one minus the sum of squared allele frequencies. Let p_i ($i = 1, 2, \ldots$) represent the frequency of allele *i*; then the heterozygosity of this locus is given by

$$H = 1 - \sum_{i=1}^{k} p_i^2,$$

where *H* stands for heterozygosity, and *k* is the total number of alleles present.

Here, in the total population T (the MRSA population of all hospitals involved) we calculate the heterozygosity of this total population (H_T), using

$$H_T = 1 - \sum_{i=1}^{k} p_{Ti}^2$$

with each allele's frequency (p_{T_i}) and the number of different alleles (*k*) in the total population. H_S, the heterozygosity of subpopulations (individual hospitals) is calculated similarly to H_T, except that first, we do the calculation restricted to each subpopulation and then calculate H_S as the average of all of the computed subpopulation heterozygosities,

$$H_S = \frac{1}{n} \sum_{i=1}^{n} H_{Si}$$

$$H_{Si} = 1 - \sum_{j=1}^{k_i} p_{ij}^2,$$

where *n* is the total number of subpopulations, H_{Si} is the heterozygosity of subpopulation *i*, k_i is the number of different alleles in subpopulation *i*, and p_{ij} is allele *j*'s frequency in subpopulation *i*.

Between these two levels, another level, the regional heterozygosity H_R (in the current study, a region means a group of hospitals we classified), is also calculated similarly: We take a group of hospitals and compute the heterozygosity of that group. Afterward, H_R is just the weighted average of all these regional heterozygosities, with the number of hospitals in each group being the weights. It can be shown that $H_T \geq H_R \geq H_S$ (equal signs hold when there is no population substructure). Then, the F statistics we used are defined as

$$F_{SR} = \frac{H_R - H_S}{H_R}$$

$$F_{RT} = \frac{H_T - H_R}{H_T}$$

$$F_{ST} = \frac{H_T - H_S}{H_T}.$$

Measurement of MRSA Population Similarity—Pairwise Analysis of Hospitals.
Potential individual predictors. To assess patient demographic factors that might account for genetic similarity of MRSA found in pairs of hospitals, we defined the following dichotomous variables. For dichotomized proportions, the subscript zero indicates a proportion less than or equal to the break point. Hospitals were classified for whether they were or had the following:

Over 35% of patients privately insured
Over 10% of patients on Medicaid
Over 20% of patients Hispanic
Over 10,000 admissions per year
Pediatric hospital (vs. adult).

Pairs of hospitals were classified as 00, 01/10, or 11 on each of these variables, and genetic similarity was assessed for hospitals that were similar on these variables (00 or 11) compared with pairs containing hospitals that were different (01/10).

The predictor of primary interest was patient flow. Using previously published data (17) on the number of times any patient was transferred between two hospitals (including possible multiple transfers of the same patient or discharge from the first before admission to the second, with an intervening stay at home), the flow of patients from hospital A to hospital B, T_{AB}, was defined as the proportion of hospital B's patients in a year who had a previous stay in hospital A during the year. The average flow between hospital A and B was then defined as $(T_{AB} + T_{BA})/2$ and was used in our analyses. A more detailed definition can be found in *SI Text*.
Linear regression analysis. We used multivariate linear regression to assess the predictors of genetic similarity between the MRSA populations in pairs of hospitals and used univariate regression for each single variable as a supplement. The response variable that measures heterogeneity was pairwise F_{SR}—the reduction in heterozygosity when two hospitals are viewed as a whole. We log-transformed these two variables to obtain normally distributed data. As some of the hospital pairs have flow = 0, we added 0.00001 (~50% of the smallest available data) to all flow data to perform the log transformation. To account for other possible predictors, we adjusted for the demographic variables described in the previous section by also using two indicator variables for each demographic variable. Standard regression P values do not account for the dependence among the observations induced by the fact that the response variables are genetic "distances" between pairs of hospitals. To adjust for this nonindependence, we performed multiple Mantel permutation tests (univariate Mantel test for univariate analyses) to generate permutation P values for all regression analyses and referred to these P values as permutation P values, as described elsewhere (31). Briefly, we constructed a genetic distance matrix for our response variable—pairwise F_{SR} from our data, with each element in the matrix—and d_{ij} corresponds to the log-F_{SR} of hospitals i and j and comes from the row of data that records the pairwise information of these two hospitals (i.e., patient flow, distances, etc.). Then we shuffled this matrix by each hospital—in other words, we shuffled the rows and columns in the same way 5,000 times, and the resulting matrices were flattened and paired back with predictor variables to conduct 5,000 regressions. Two-tailed P values were calculated from the distribution of t statistics of corresponding coefficients generated.

In addition, for each pair of hospitals, we also adjusted for (*i*) sample size, by using the average number of *spa*-typed isolates of the two hospitals, and (*ii*) distance between the two hospitals, calculated on the basis of their longitudinal and latitudinal data, in kilometers.

Univariate plots of the response vs. individual predictors were checked visually for outliers.

Measurement of MRSA Population Similarity—Analysis of Groups of Hospitals. As a complementary approach, we considered whether grouping the hospitals into a small number of groups on the basis of the demographic characteristics used above of their patient populations or, of more direct interest, on the basis of their patterns of patient sharing, would create groups that captured some of the population genetic structure of the MRSA in the hospitals. To assess this possibility, we sought both to assess the extent to which the best possible grouping could create groups that are genetically homogeneous (the value of F_{RT} obtained by an optimal grouping) and to assess how much genetic structure would be captured in randomly constructed groupings of all 30 hospitals (the range of F_{RT} values obtained by random groupings). The first assessment was done by using a genetic algorithm (GA) to give an approximate numerical value because an exhaustive/exact method is computationally infeasible, and the second assessment was done by creating groupings in which hospitals were randomly assigned to group membership.
GA—Evaluation of grouping efficiency. To establish a standard for the possible extent to which any grouping scheme of hospitals could define genetically similar MRSA populations ("grouping scheme" or "grouping" refers to a specific group assignment of all hospitals), we attempted to find the groupings of hospitals (using no information about the hospitals themselves) that maximized F_{RT}—the measure of genetic heterogeneity between groups—by using a GA. We implemented this method by generating random groupings at a given number of groups and evaluating F_{RT} for each grouping and then evolving the groupings by enriching and combining groupings with high F_{RT}. The goal was to approximate the optimal groupings at a given number of groups to serve as a reference of maximum F_{RT}. This method was repeated for groupings with two, four, or six groups (we primarily used two and four groups, and the result of six groups was used to show the trend of F_{RT} when the number of groups increases). These groupings are referred to as GA groupings. Details of this algorithm can be found in *SI Text* and Fig. S2.
Evaluation of null distribution of F_{RT}. To evaluate whether groupings of hospitals based on any given measurement contain information about the genetic structure of the MRSA population (F_{RT}) greater than expected by chance alone, we created a null distribution of F_{RT} for randomly chosen groupings that divided hospitals into two, four, and six groups. For each given number of groups, we randomly generated 15,000 groupings and tabulated the distribution of F_{RT} from these results (these groupings are referred to as random groupings of k groups). We obtained the mean and 95% coverage interval of F_{RT}s for these random groupings. The P value (double sided) of a given F_{RT} for a particular grouping is defined by the proportion of randomly generated F_{RT}s that are of equal distance or farther away from the mean than it is.
Grouping of hospitals by prespecified categories. We first grouped all hospitals using categorizing variables specified in *Measurement of MRSA Population Similarity—Pairwise Analysis of Hospitals* , including private insurance proportion, Medicaid coverage, size (annual admission), ethnicity (Hispanic), and pediatric vs. adult-only hospitals (the groups are named by their grouping criteria).
Grouping of hospitals by the information theoretic approach (network grouping). For patient sharing, we used the algorithm of ref. 23 to identify "modules" or neighborhoods within the network of hospitals on the basis of patient flow. To do so, we considered each hospital as a "node" in the network and between each hospital constructed an undirected edge representing patient sharing with weight determined by the proportion of shared patients, as given in *Measurement of MRSA Population Similarity—Pairwise Analysis of Hospitals*. The algorithm defines neighborhoods—roughly speaking—as sets of nodes in which an imaginary random walker, traversing edges of the network with probabilities proportional to the edge weight, would be much more likely to stay within a set. Thus, sets of nodes that are well connected with one another will tend to be in the same group; in our case, sets of hospitals that share many patients with one another will tend to be in the same group, and sets of hospitals that do not share many patients with one another will tend to be in different groups. Technical details of this method are given in *SI Text*, which summarizes the account given in ref. 23. This method is referred to as network grouping.

ACKNOWLEDGMENTS. This work was supported by Models of Infectious Disease Agent Study Award U54GM088558 (to W.K. and M.L.), by US National Institute Of General Medical Sciences Grant U01 GM76672 (to S.S.H. and T.R.A.),

by funds from the University of California Irvine School of Medicine (to S.S.H., K. R.E., C.C.N., and C.R.M.), by Award WT089472MA from the UK Wellcome Trust (to B.G.S.), and by a UK Biotechnology and Biological Sciences Research Council Doctoral Training Grant (to L.O.H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

1. Cooper BS, et al. (2004) Methicillin-resistant Staphylococcus aureus in hospitals and the community: Stealth dynamics and control catastrophes. *Proc Natl Acad Sci USA* 101:10223–10228.
2. Donker T, Wallinga J, Grundmann H (2010) Patient referral patterns and the spread of hospital-acquired infections through national health care networks. *PLoS Comput Biol* 6:e1000715.
3. Grundmann H, et al.; European Staphylococcal Reference Laboratory Working Group (2010) Geographic distribution of Staphylococcus aureus causing invasive infections in Europe: A molecular-epidemiological analysis. *PLoS Med* 7:e1000215.
4. Feil EJ, et al. (2003) How clonal is Staphylococcus aureus? *J Bacteriol* 185:3307–3316.
5. Grundmann H, Aires-de-Sousa M, Boyce J, Tiemersma E (2006) Emergence and resurgence of meticillin-resistant Staphylococcus aureus as a public-health threat. *Lancet* 368:874–885.
6. Livermore DM, Pearson A (2007) Antibiotic resistance: Location, location, location. *Clin Microbiol Infect* 13(Suppl 2):7–16.
7. Tiemersma EW, et al.; European Antimicrobial Resistance Surveillance System Participants (2004) Methicillin-resistant staphylococcus aureus in Europe, 1999–2002. *Emerg Infect Dis* 10:1627–1634.
8. UK Health Protection Agency (2008) Surveillance of healthcare associated infections report. Technical report. Available at http://www.hpa.org.uk/web/HPAwebFile/HPAweb_C/1216193833496. Accessed February 28, 2012.
9. Dominguez MA, de Lencastre H, Linares J, Tomasz A (1994) Spread and maintenance of a dominant methicillin-resistant Staphylococcus aureus (MRSA) clone during an outbreak of MRSA disease in a Spanish hospital. *J Clin Microbiol* 32:2081–2087.
10. Harbarth S, et al. (2000) Effect of delayed infection control measures on a hospital outbreak of methicillin-resistant Staphylococcus aureus. *J Hosp Infect* 46:43–49.
11. Kotilainen P, et al. (2003) Elimination of epidemic methicillin-resistant Staphylococcus aureus from a university hospital and district institutions, Finland. *Emerg Infect Dis* 9: 169–175.
12. Roman RS, et al. (1997) Rapid geographic spread of a methicillin-resistant Staphylococcus aureus strain. *Clin Infect Dis* 25:698–705.
13. UK Department of Health (2005) Mrsa surveillance system: Results. Technical report. Available at http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsStatistics/DH_4085951. Accessed August 2011.
14. Scanvic A, et al. (2001) Duration of colonization by methicillin-resistant Staphylococcus aureus after hospital discharge and risk factors for prolonged carriage. *Clin Infect Dis* 32:1393–1398.
15. Robotham JV, Scarff CA, Jenkins DR, Medley GF (2007) Meticillin-resistant Staphylococcus aureus (MRSA) in hospitals and the community: Model predictions based on the UK situation. *J Hosp Infect* 65(Suppl 2):93–99.
16. Smith DL, Levin SA, Laxminarayan R (2005) Strategic interactions in multi-institutional epidemics of antibiotic resistance. *Proc Natl Acad Sci USA* 102:3153–3158.
17. Huang SS, et al. (2010) Quantifying interhospital patient sharing as a mechanism for infectious disease spread. *Infect Control Hosp Epidemiol* 31(11):1160–1169.
18. Roberts RB, et al.; MRSA Collaborative Study Group (1998) Molecular epidemiology of methicillin-resistant Staphylococcus aureus in 12 New York hospitals. *J Infect Dis* 178: 164–171.
19. Kho AN, Lemmon L, Commiskey M, Wilson SJ, McDonald CJ (2008) Use of a regional health information exchange to detect crossover of patients with MRSA between urban hospitals. *J Am Med Inform Assoc* 15:212–216.
20. Frénay HM, et al. (1996) Molecular typing of methicillin-resistant Staphylococcus aureus on the basis of protein A gene polymorphism. *Eur J Clin Microbiol Infect Dis* 15:60–64.
21. Mellmann A, et al. (2006) Automated DNA sequence-based early warning system for the detection of methicillin-resistant Staphylococcus aureus outbreaks. *PLoS Med* 3: e33.
22. Harmsen D, et al. (2003) Typing of methicillin-resistant Staphylococcus aureus in a university hospital setting by using novel software for spa repeat determination and database management. *J Clin Microbiol* 41:5442–5448.
23. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci USA* 105:1118–1123.
24. Lee BY, et al. (2011) Social network analysis of patient sharing among hospitals in Orange County, California. *Am J Public Health* 101:707–713.
25. Harris SR, et al. (2010) Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327:469–474.
26. Ko KS, et al. (2006) Molecular characterization of methicillin-resistant Staphylococcus aureus spread by neonates transferred from primary obstetrics clinics to a tertiary care hospital in Korea. *Infect Control Hosp Epidemiol* 27:593–597.
27. Lee BY, et al. (2011) Modeling the spread of methicillin-resistant Staphylococcus aureus (MRSA) outbreaks throughout the hospitals in Orange County, California. *Infect Control Hosp Epidemiol* 32:562–572.
28. Outterson K, Yevtukhova O (2011) Germ shed management in the United States. *Antibiotic Policies: Controlling Hospital-Associated Infection*, eds Gould IM, van der Meer J (Springer, Berlin), pp 163–182.
29. Hudson LO, et al. (2012) Differences in methicillin-resistant *Staphylococcus aureus* strains isolated from pediatric and adult patients from hospitals in a large county in California. *J Clin Microbiol* 50(3):573–579.
30. Hartl DL, Clark AG (1997) *Principles of Population Genetics* (Sinauer, Sunderland, MA), 3rd Ed.
31. Manly BFJ (2007) *Randomization, Bootstrap and Monte Carlo Methods in Biology* (Chapman & Hall/CRC).